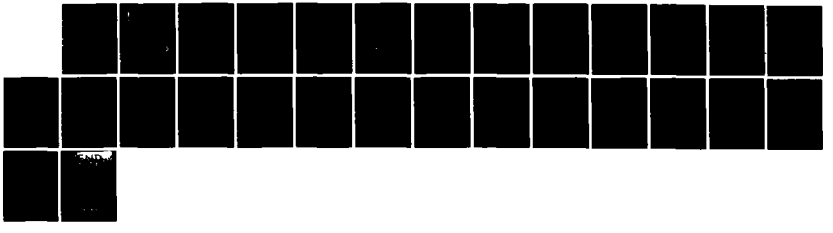
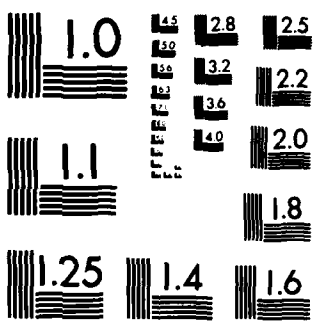
AD-A146 878 LITERATURE REVIEW OF VOICE RECOGNITION AND GENERATION TECHNOLOGY FOR ARMY HELICOPTER APPLICATIONS(U) HUMAN ENGINEERING LAB ABERDEEN PROVING GROUND MD K A CHRIST

1/1

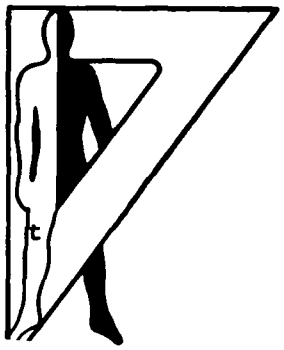
UNCLASSIFIED AUG 84 HEL-TN-11-84 F/G 17/2 NL






OPY RESOLUTION TEST CHART

(12)



AD

AD-A146 878

Technical Note 11-84

# LITERATURE REVIEW OF VOICE RECOGNITION AND GENERATION TECHNOLOGY FOR ARMY HELICOPTER APPLICATIONS

Kathleen A. Christ



August 1984



U. S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial products.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER: Technical Note 11-84 | 2. GOVT ACCESSION NO.: AD-A146 878 |
| 3. RECIPIENT'S CATALOG NUMBER | |
| 4. TITLE *(and Subtitle)*: LITERATURE REVIEW OF VOICE RECOGNITION AND GENERATION TECHNOLOGY FOR ARMY HELICOPTER APPLICATIONS | 5. TYPE OF REPORT & PERIOD COVERED: Final Report |
| | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s): Kathleen A. Christ | 8. CONTRACT OR GRANT NUMBER(s) |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: US Army Human Engineering Laboratory, Aberdeen Proving Ground, MD 21005-5001 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS | 12. REPORT DATE: August 1984 |
| | 13. NUMBER OF PAGES: 24 |
| 14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* | 15. SECURITY CLASS. *(of this report)*: Unclassified |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

| | |
|---|---|
| Voice Recognition | Stress |
| Voice Generation | Noise |
| Voice Data Entry | Human Factors |
| Manual Data Entry | Voice Warning Systems |

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report is a literature review on the topics of voice recognition and generation. Areas covered are manual versus vocal data input, vocabulary, stress and workload, noise, protective masks, feedback, and voice warning systems.

Results of the studies presented in this report indicate that voice data entry has less of an impact on a pilot's flight performance, during low-level flying and other difficult missions, than manual data entry.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

However, the stress resulting from such missions may cause the pilot's voice to change, reducing the recognition accuracy of the system. The noise present in helicopter cockpits also causes the recognition accuracy to decrease. Noise-cancelling devices are being developed and improved upon to increase the recognition performance in noisy environments.

Future research in the fields of voice recognition and generation should be conducted in the areas of stress and workload, vocabulary, and the types of voice generation best suited for the helicopter cockpit. Also, specific tasks should be studied to determine whether voice recognition and generation can be effectively applied.

Technical Note 11-84

# LITERATURE REVIEW OF VOICE RECOGNITION AND GENERATION TECHNOLOGY FOR ARMY HELICOPTER APPLICATIONS

Kathleen A. Christ

August 1984

APPROVED:

JOHN D. WEISZ
Director
US Army Human Engineering Laboratory

US ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland

CONTENTS

# LITERATURE REVIEW OF VOICE RECOGNITION AND GENERATION

# TECHNOLOGY FOR ARMY HELICOPTER APPLICATIONS

## INTRODUCTION

The US Army Human Engineering Laboratory (USAHEL) is currently investigating the use of voice recognition and generation in helicopter cockpits. The application of voice technology is intended to decrease the workload of the pilot. Enabled to enter and request data by voice, the pilot's hands are free to maintain contact with the flight controls.

This report is a literature review on the topics of voice recognition and generation. Information was obtained through the Defense Technical Information Center (DTIC), Human Factors Index, conference proceedings, books and journals. Because of the improvements constantly being made in voice technology, only sources since 1978 were used. Also, only the sources which dealt with topics pertaining to helicopter cockpits were consulted. Unless otherwise stated, the results of the studies presented in this report were determined to be statistically significant by the respective authors.

The objectives of this review were to summarize the effectiveness and reliability of present voice recognition and generation systems and to determine in what areas future research should be initated.

### Voice Recognition and Speech Generation

There are four types of voice recognition: speaker dependent, speaker independent, isolated word, and continuous speech. Depending upon the application, each voice recognition system has advantages and disadvantages.

Speaker dependency means that speakers must train the system to their own voice. Voice templates are formed by repeating the vocabulary several times before actual use. When the recognition system is in actual use, the spoken utterances are compared to the templates to obtain recognition. A few systems have gotten around the tedious process of training the system. In such systems, one prior training pass is required before use. Then the templates can be updated during usage if any recognition errors occur. Speaker dependent systems afford some degree of security because only those who have trained their voice to the system can gain access.

Speaker independent systems allow any speaker to access control of the machine by voice, without providing sample templates of their voice. Recognition is based upon voice templates formed from a selected sampling of a variety of different speakers. Such systems can be useful in telecommunications. So far, however, the vocabulary usable on such systems is limited to numbers and a few words. Where security is a primary concern, speaker independent systems should not be considered.

Isolated word recognizers require the user to pause between each utterance. This slows data entry and also creates an unnatural way of speaking. Speaker dependent, isolated word machines were the first to be developed. They can now handle large vocabularies (200-300 words) and obtain recognition accuracies of more than 99 percent. The research reviewed in the following sections uses an isolated word, speaker dependent recognizer.

In the past few years, continuous speech recognizers have developed to the point where they can now obtain recognition accuracies between 55 and 97 percent (Aretz, 1983). Continuous speech enables the user to enter data in a more natural speaking manner. No pause is required between utterances. However, the problem in developing such systems is determining when one word ends and the next begins.

## Speech Generation

There are two types of speech generation: digitized speech systems and synthesized speech systems. Digitized speech systems store the speech wave form for each word as a binary value. When a word is played back, it is converted from its digitized form back into its original spoken form. Synthesized speech models human vocal responses electronically. The phonemes which make up human speech are electronically reproduced to form speech-like quality. The prior recording and storing of human utterances are not required (North & Lea, 1982).

The perception of synthetic speech is not as accurate as that of natural speech (Nusbaum, Schwab, & Pisoni, 1983). This difficulty may possibly be overcome by training the listener to the synthesized speech. Digitized speech, because it stores human utterances, can be understood as well as natural speech.

## Advantages and Disadvantages of Voice Technology

Beek and Vonusa (1983) listed several advantages and disadvantages of voice recognition and generation. In the sections which follow, these points will be discussed in more detail. Some of the advantages are:

a. Can be faster than other modes of communication.

b. Can be more accurate than other modes of communication.

c. Most natural form of communication.

d. Can reduce visual and motor workload.

e. Increases in value proportional to complexity of information being processed.

f. Requires less effort and motor activity than other communication modes.

g. Frees hands and eyes (does not require physical contact with voice recognizer).

h. Can be operated in darkened environments.

i. Does not require direct line of sight to the voice recognizer.

j. Allows for more operator mobility than conventional data input and output device.

Some of the disadvantages are:

a. Competing acoustic sources may interfere.

b. Variety of physical conditions can change acoustic characteristics of speech.

c. Fatigue may change speech characteristics.

d. Physical ailments may change speech characteristics.

e. Speech not private, may be overheard.

f. Psychological changes (stress) in speaker may change speech characteristics.

g. Speech generation may interfere with other auditory indicators.

h. Speech generation can be slower than visual displays. The user must wait for each word to be generated before proceeding, whereas visual displays enable the user to scan the information presented for the relevant data only.

However, Beek and Vonusa stress that it is necessary to distinguish between the disadvantages of speech communication and the limits of current speech technology. The limitations of the technology may be overcome by future voice systems, whereas the disadvantages of speech communication are a result of the environmental factors affecting the voice system.

## VOICE RECOGNITION

### Speaker Dependent Systems Used as Speaker Independent

The Naval Postgraduate School performed a study to determine the effectiveness of using speaker dependent systems as speaker independent systems (Poock, Schwalm, Roland, & Martin, 1982). Recognition accuracy was compared under the following conditions:

1. When only the training pattern of the speaker is used as a template reference.

2. When the training patterns of the speaker as well as four other speakers are used as a template reference.

3. When only the patterns of the four other speakers are used as a template reference (speaker independent mode).

Results of this study showed that an accuracy level of 99 percent can be attained under condition 2, while under condition 3, a 95 percent accuracy level can be attained, with no significant increase in misrecognitions. Breaking down the errors which occurred into nonrecognitions (system did not recognize input) and misrecognitions (system recognized wrong input), the misrecognitions were not affected under any of the conditions. The nonrecognitions increased slightly (4 percent) under condition 3, but were not affected under condition 2. Although no objective analysis was performed, the authors noted a similarity in the subjects' voices. Also, the author suggested that the experience level of the users may have had an affect on the results. In this study, all subjects had experience training and testing the system. They recommended that future research should be conducted to determine the accuracy level attained by a speaker who has not trained the system.

### Manual Versus Vocal Input

Currently, data entry in aircraft is accomplished by using a keyboard. This method of data entry requires the pilots to remove their hands from the flight controls. During critical phases of flight, it is imperative that the pilot maintain control of the aircraft. This cannot be accomplished if the pilot must also enter required data. In the helicopter of the future, where only one pilot may be present, maintaining control of the aircraft while entering data becomes much more difficult, if not impossible. Several studies have been done to determine which mode of entry is more effective under different circumstances.

In a study performed to determine the effectiveness of voice recognition over manual entry in a time-sharing environment (Skriver, 1979), subjects were asked to perform a digit entry task concurrently with a tracking task. The experiment was performed in isolation, so no outside factors interfered with either the keyboard entry or the voice entry. The digit entry task required the subjects to enter a digit corresponding to the digit displayed on a cathode-ray tube (CRT). The tracking task involved maintaining the position of a cursor on the center of a horizontal track displayed on the CRT. Control of the cursor was through a hand controller. Subjects were given feedback concerning their performance.

Results of the study indicated that while both manual and vocal performance declined during the tracking task, fewer errors occurred with voice data entry. The performance of the tracking task also declined during the dual task mode. But as before, fewer input errors occurred during the voice data entry mode.

Another study was performed at the Naval Postgraduate School (Ruess, 1982) comparing voice recognition and keyboard entry devices. Both systems were analyzed to determine time to load, input and output accuracies, and time versus accuracy in retargeting air launch cruise missiles. Results of this study indicated that keyboard entry was better than voice in time to load and input accuracies. Fewer corrections were also made during the keyboard entry. As the time to complete the task increased, the accuracy also increased. However, despite the slower input speeds, 60 percent of the subjects preferred voice entry over keyboard.

As in all the studies presented in this report, an isolated word recognizer was used for the voice data entry. The author concluded that a connected speech recognizer would have been more appropriate, allowing faster input. Ruess recommended that for tasks requiring quick loading of single characters, continuous speech recognizers should be compared to keyboard entry devices.

Another study comparing voice and manual data entry was conducted using a fighter cockpit simulator (Aretz, 1983). Both modes were compared in a single and dual task condition based on flight performance, response time, and errors. Navigation and weapons tasks were performed using vocal and manual modes. In the single task condition, only the multifunction control was operated by the pilot (manually or vocally). The dual task condition required the pilot to fly the simulator in addition to operating the control. The subjects used for this study were operationally qualified Air Force pilots.

When data entry was the primary task, it was concluded that the manual data entry was the most effective. The time to complete the data entry was shorter for the manual mode than for the voice mode. However, if flying the aircraft was the primary task, then it was concluded that the vocal response mode had the least impact on flight performance and was therefore more effective. Tracking performance did not change for the vocal response mode in the single and dual task mode, but declined for the manual response mode under the dual task mode.

General Dynamics Corporation performed a study using a cockpit mock-up (Wyatt, 1983). Radar, weapons, and flight-control data were controlled vocally instead of manually. This study corresponded with the joint Air Force, Navy, and NASA advanced fighter technology integrator (AFTI/F-16) program. The pilot was required to fly in formation at a constant airspeed and to fire weapons at a specified target. The comparison of vocal versus manual data entry occurred during additional tasks assigned to the pilot.

At the end of the study, the pilots agreed that their workload decreased when they performed their tasks by voice. They stated that the voice mode allowed more time for crosschecking instruments, selecting targets, and firing weapons. Overall, voice entry was more effective than manual entry.

Although the study by the Naval Postgraduate School (Ruess, 1982) seemed to contrast with the other studies discussed, they all concluded that voice data entry can be more effective than manual data entry in certain situations. A few other studies similar to those mentioned also support up this conclusion (Jay, 1981; Coler, 1983). Based upon the research presented, the following statements can be made:

a. When flight control is critical, data entry by voice is less disruptive and more effective than keyboard entry.

b. Voice data entry is preferred by users over manual data entry in certain situations, especially when the workload is excessively heavy.

c. Both voice and manual systems should be provided, with neither completely replacing the other.

## Vocabulary Size

The selection and size of a vocabulary can have an effect on the recognition performance of a voice system, as well as the performance of the user. Pilots may have trouble remembering larger vocabularies when they are faced with a multitude of other tasks to be performed (Wyatt, 1983). Larger vocabularies also require longer training passes and larger storage space.

As far as single vocabulary words, long words are more easily recognized than short words. Also, the phonetic version of the alphabet (alpha, bravo, charlie, etc.) causes less confusion than the orthographic (A, B, C, etc.) version (Bridle, 1983).

The Naval Postgraduate School examined voice recognition performance under varying vocabulary sizes (Poock, 1981). Using vocabulary sizes ranging from 20 to 240 words, subjects trained and then tested their voice patterns. Results indicated that the error rate did not increase significantly as a function of vocabulary size. However, it should be pointed out that the subjects were not required to perform any other tasks aside from repeating the trained vocabulary.

Selection of the vocabulary words can help alleviate some of the problems caused by the size of the vocabulary. Words should be selected that are familiar to the user and correspond to the system. The careful selection of words will aid the user in remembering the vocabulary. Words which sound similar, or may have a similar meaning, should be avoided.

The TALK and TYPE system has been shown to be more efficient than typing when large vocabularies are involved (Welch & Shamsi, 1980). These systems require the user to type the first letter of a word as it is spoken into the voice recognition systems. The active vocabulary is reduced to only those words beginning with the typed letter. Recognition accuracy is thus increased.

The Naval Air Development Center is currently working on an improved syntax structure for airborne applications. The system would enable the pilots to communicate in a more natural form.

The system is an extension of the voice recognition and synthesis (VRAS) system developed earlier (Stokes & Dow, 1980). VRAS enables users to speak commands in sentences rather than in individual words. A vocabulary between 100 to 300 words is defined to meet a specific task-processing situation. The syntactical properties of the vocabulary are also defined. Thus, the computer expects certain words in a given situation. This enables the recognizer to differentiate between homonyms (such as two and to) in most situations.

Vocabulary consideration is just one step in the analysis and implementation of a voice recognition system. More work dealing with the improvement of syntax structures needs to be accomplished in order for voice recognition to be truly beneficial. Along with improved syntax structures, the vocabulary should be carefully selected to minimize size and eliminate confusion of similar words.

## Stress and Mental Loading

As stated on page 4, voice recognition systems can obtain recognition accuracies of more than 99 precent. Factors like emotional and physical stress can affect a person's voice, thus affecting recognition performance (Armstrong, 1980). Recognizers must be able to adapt to the changes in a user's voice. The differences between normal speech and speech under stressful conditions are as follows:

a. Difference in volume, in both directions (soft,loud)

b. Difference in fundamental frequency (the frequency of vibration of the vocal cords), in both directions (high,loud) - in one situation, fundamental frequency may tend to increase from beginning to end of utterance. In another situation, frequency may tend to decrease from beginning to end of same utterance.

c. Difference in precision of articulation - under stress, a person may slur syllables together or omit speech sounds altogether.

d. Difference in monotonicity of speech - under stress, a person may speak in a monotone.

Stress is a difficult condition to induce in a laboratory setting. Hogan and Hogan (1982) determined that psychological stressors in laboratory studies depend on the following:

a. Prior experience with a class of stressors - if one lacks experience with the stressors, then no stress response will occur.

b. Remembering the experience - if one does not recall that experience (represses the memory), then no stress response will occur.

c. Recognizing the present stimulus is an instance of the earlier class of stressors - if one does not recognize that the present conditions are similar to past threatening circumstances (and other types of repressors involving redirection of attention), then no stress response will occur.

d. Believing the likelihood of threatened occurrence is above some subjective threshold - if one does not believe there is a significant likelihood of the threat materializing, then no stress response will occur.

When developing tasks to induce stress, these elements should be considered to attain the desired level of stress.

The Naval Postgraduate School performed a preliminary inquiry into the effects of stress (French, 1983). Subjects were placed under time-induced stress. Both experienced and inexperienced users of voice technology were selected as subjects.

The study consisted of three phases. In the first phase, subjects completed the task at their own pace. The experimenter advised the subjects not to rush, but also not to linger. The time to complete the task was recorded for each subject and used as the baseline time for that person. The next two phases of the study required the subjects to complete the task first within two-thirds of their baseline time, then within one-third of the time. Subjects were informed of the amount of time given to complete task, and a timer was visible to the subjects.

For the experienced users, recognition rates decreased from 99.5 percent to 89.7 percent as the time to complete the task was reduced. However, for the inexperienced user, recognition rates increased from 90.7 percent to 94.1 percent, but for phase 3, decreased to 91 percent. Overall, when all users were analyzed, it was concluded that time-induced stress has a negative effect on recognition accuracy.

The stress induced on by operator mental loading was the basis for two other experiments at the Naval Postgraduate School (Armstrong & Poock, 1981a; Armstrong & Poock, 1981b). Both studies utilized a General Dynamics response analysis tester (RATER) to induce mental loading.

For the first study, three different RATER tasks were required. Four symbols (triangle, circle, cross, and diamond) were randomly displayed at a constant rate of one symbol every 1.5 seconds. A response button corresponding to each of the symbols was located on the console. The subjects were placed under four different mental loading conditions:

a. No RATER Task - subject operated only voice recognition device.

b. RATER delay zero - subject responded with symbol which corresponded to symbol being displayed.

c. RATER delay one - subject responded with symbol which appeared the previous trial.

d. RATER delay two - subject responded with symbol which appeared two trials earlier.

For the voice task, the subjects were given a 2.5-minute tape recording of 50 words arranged in random order. The words were presented at a constant rate of one every 3 seconds. The subjects were instructed that the highest priority task was to repeat the vocabulary words one at a time as they appeared.

It was shown that the recognition error rates in the three RATER tasks were 23% greater than the error rate of the no RATER task condition. Thus, operator mental loading has a significant differential effect on recognition error rate.

During the first study, it was also found that performance during the first 2-1/2 minutes of the experiments differed from the second 2-1/2 minutes. This led to the second study to determine what effect the task duration had on recognition accuracy. Only two experimental conditions were used: no RATER task and RATER delay one (as defined earlier) task. Subjects were not given feedback and did not know how much time remained to complete the task.

The results of the second study revealed that the error rate was higher for the RATER task condition than for the no RATER task condition. The length of time to complete the task did not affect error rates.

Based on the above studies, the following conclusions can be drawn:

a. Mental loading, at any level, decreases recognition accuracy.

b. Subject error rate increased when mental loading increased.

c. Task duration does not affect subjects' error rate.

More research dealing with stress and mental loading is needed to resolve problems such as the effect of time-induced stress. The only literature available on the effects of stress on voice recognition has come from the Naval Postgraduate School. Methods other than increased mental workload must be developed to induce stress. Further research dealing with the changes the human voice undergoes as a result of stress is also needed.

## Noise

The environment of an aircraft cockpit presents another problem when considering the application of voice technology--noise. Sustained levels of ambient noise of up to 115 decibels (dB) must be overcome by the voice recognizer (Coler, 1983).

The effects of noise on voice recognition were the topic of a study performed at the Naval Postgraduate School (Elster, 1980). Three levels of noise (38, 65, and 75 dBA, where dBA refers to the A-weighting network which estimates the interference of noise upon speech) were used. The 38 dBA white noise was considered ambient noise, and the 65 and 75 dBA were conversational noise. The subjects trained their voice in one level of noise but tested in all three noise levels.

The results indicated that if the system was trained at 38, 65, or 75 dBA, then the performance of the system would be satisfactory when used in a 38 or 65 dBA noise environment. However, to be used in a 75 dBA noise environment, the system should be trained in a 65 or 75 dBA noise environment. The conclusion drawn from this experiment indicated that only the noise condition during testing affected the performance of the voice recognition system.

In a study conducted at the US Army Avionics Research and Development Activity (AVRADA), noise levels of higher intensity were used. Subjects tested a voice recognition system in three environments: no helicopter noise, 103 dBA, and 107 dBA. The noise environments are those found in a UH-60 helicopter.

The highest recognizer accuracy achieved was found when the vocabulary was trained in actual noise environment. This is basically the same results obtained by Elster. AVRADA has also developed noise cancelling devices to help deal with the noise present in a cockpit (Reed, 1982).

An experiment at NASA-Ames evaluated a voice recognition device in noise when the subject was also required to perform a tracking task and enter data (Coler, 1983). Performance was evaluated for three different conditions: no noise, 90 dBA helicopter, and 100 dBA helicopter noise.

Results of this study showed that recognition accuracies declined from 99 percent when there was no noise to 97 percent when 100 dBA of noise was present. However, when the training of the voice recognizer and the testing under the same noise levels, recognition accuracy at the higher noises increases to 99%.

In the General Dynamics study dealing with the AFTI/F16 aircraft (Wyatt, 1983), the effect of noise was also evaluated. Although initial testing produced performance levels in the range of 82 percent accuracy at 85 dB and 13 percent accuracy at 115 dB, a militarized version of a voice recognition system, capable of withstanding the vibration and motion of an aircraft, improved performance. Improved performance was obtained when the system was trained at 85 dB and operated at 115 dB.

According to the results of the studies presented here, the effect of noise on the performance of a voice recognition system can be resolved partially by training in a similar environment. Noise-cancelling devices can also aid in elimination of the effects of noise.

Protective Masks

Situations may arise when it may become necessary to operate voice recognition equipment while wearing a protective mask. In areas where a number of voice recognition users are present, stenographer's masks may be worn to prevent interference between talkers. In the military, the possibility of chemical warfare creates the need for personnel to wear protective masks. The use of masks does have an impact on the recognition accuracy of voice recognition equipment.

Results of a stenographer's mask study performed at the Naval Postgraduate School showed that the masks did cause an increase in the misrecognition rate (Poock, Schwalm, & Roland, 1982). When no mask was worn, an accuracy rate of 98.2 percent was achieved. Under the masked condition, an average accuracy rate of 94.7 percent was attained. It was determined that the level of experience the user had in using the mask and microphone affected the error rate. The more experience the user had, the lower the impact on the misrecognition rate.

A second experiment was performed to determine whether a more restrictive protective mask had a greater effect on recognition accuracy. (Poock, Roland & Schwalm, 1983). The Army M24 field protective mask was used. The microphone in the M24 is mounted internally and placed directly in front of the mouth, below the lower lip.

The results showed that although experience did improve the misrecognition rate, it was still higher than the error rate recorded for the stenographer's mask. The authors listed some possible causes for the increased error rate when using gas masks:

a. Front-mounted microphone placed too close to and directly in front of mouth. The researchers believe this is the worst position for the microphone.

b. Noise occurred at beginning and ending of words as a result of the breathing hose being placed next to the microphone. Users took a breath before and after utterance, creating a noise.

c. The masks may not have always been adjusted properly. This would make it harder to breathe comfortably.

d. Users' attitudes may have had an effect. Users became frustrated at the operation of the voice recognizer. Also, the gas masks are uncomfortable to wear.

The US Army Human Engineering Laboratory performed a similar experiment to those conducted by Poock and his associates (Malkin, 1983). Two different protective masks were used: M24 aviator mask and the developmental XM33 aviator mask. The location of the microphone in the M24 aviator's mask is similar to that in the M24 field mask. The microphone on the XM33 is located outside the mask and behind a diaphragm.

The results obtained from this experiment support those of the Naval Postgraduate School. Experience in the use of the masks and microphone will alleviate some of the recognition errors. Of the two masks, the M24 produced better results than the XM33. However, the results were still below the recognition accuracy level supposedly obtainable by the voice recognition system.

If voice recognition is to be used in situations where the need to wear protective masks may arise, improvements need to be made in their use. The problems caused by the effects of breathing sounds inside the mask and the placement of the microphone need to be resolved.

## Feedback

Feedback of information entered through a voice recognition system can be accomplished by two methods: visual and aural. Visual feedback is simply the display of input data on a CRT. Aural feedback utilizes voice generation to relay information from the recognition system to the user.

Groner and Gilblom (1983) listed a few disadvantages of the use of visual feedback:

a. User must remain within a certain area to keep display in viewing range at all times.

b. User must discern between instructions and data which may be displayed in the same field.

c. User must divert eyes from task he is performing causing inefficiency and eye fatigue.

d. There is no guarantee user will actually attempt to access important information displayed at the right time.

In airborne applications, there are other ways to present information visually without diverting the pilot's eyes from his flying task. Head-up or helmet mounted displays could be used to enable pilots to maintain visual contact with the outside environment.

In a study performed at the US Army Aeromechanics Laboratory (Voorhees, Marchionda, and Atchison, 1982), three display formats were compared to determine which format conveyed the needed information without affecting flight performance. The display formats compared were as follows:

a. Conventional dials: Three single needle dials located on a console in front and below the level of the viewing screen.

b. Head-up display: Ribbon type gauges were drawn by the graphics system on the CRT around the sides of the tracking task.

c. Auditory display: Synthesized speech feedback responses to spoken requests, also unrequested voice warning messages, occurred if the subject exceeded the limits placed on airspeed, altitude, or torque.

The information displayed by these different formats included airspeed, altitude, and torque. A 90dBA background helicopter cockpit noise was also present.

Results of this experiment showed an improvement in performance using the auditory display compared to using the other two displays. When comparing only visual displays, the head-up display improved performance over the conventional dials.

Feuge and Geer (1978) found that when the users of a voice recognition system entered data which was fed back aurally, they tended to wait for the feedback before continuing. The authors suggested that this delay may have resulted from the users waiting to see if the recognition system recognized the proper input. In any case, the pause between input and output increases the time to complete the task. In situations where speed of input is the greatest concern, aural feedback does not appear to be advantageous.

If aural feedback is to be used in an aircraft cockpit, only the necessary information should be presented. The pilot must contend with noise, monitor several radio channels, listen for warnings, and possibly communicate with other crewmembers. Manaker (1982) states that although part of a radio message may not be intelligible to the pilot, the message can still be understood. However, there is still the possibility of the pilot experiencing an information overload.

## VOICE GENERATION

### Voice Warning Systems

Voice generation is considered to be a less difficult problem than voice recognition. It is currently feasible to utilize voice generation in a number of areas. Chrysler offers an electronic voice alert system as a standard feature in several of their top-of-the-line automobiles (Finkelstein, 1983). Aircraft cockpits can present warning signals through the use of voice generation systems. However, the cockpit presents a different environment in which to apply voice generation.

When attempting to install a voice warning system in an aircraft, there are several guidelines, as listed by Berson and Associates (1981), which should be followed:

a. Purpose: Voice messages should be used when the crew must act rapidly and enable the pilot to transfer workload from the visual to the auditory channel.

b. Voice Characteristics: Select voice characteristics that are highly distinctive and intelligible.

c. Voice Inflection: Voice messages should be presented with a monotone inflection. An urgent sounding message for a warning may place additional stress on the pilot. Also, less time is taken to present a monotone warning than a conversational warning.

d. Intensity: Voice messages should be presented at an intensity level of 3-8dB above ambient noise level.

e. Onset Coordination: The time between the alerting signal and the voice message should be in the range of .15 to .50 second.

f. Message Content: Voice messages should be constructed using short pauses so that the problem or action to be taken is clearly understandable. For time-critical warnings, messages should contain two elements (action and direction). Other messages should contain three elements (general heading, location or subsystem, and nature of problem).

g. Accommodation of Multiple Voice Messages: A prioritization scheme should be used to enable the alerting system to present messages in order of criticality. The message "multiple alerts" should be presented when two or more warnings occur simultaneously, or when two or more cautions occur simultaneously without the existence of any higher priority alerts. The warnings should be conveyed as quickly and efficiently as possible. The criteria for classification of warnings and cautions is as follows:

(1) Warnings: Emergency operational or aircraft system conditions that require immediate corrective action.

(2) Cautions: Abnormal operational or aircraft system conditions that require immediate crew awareness and prompt action.

h. Message Cancellation: Time-critical voice messages should be able to be cancelled manually, or automatically, when the situation no longer exists.

The use of voice as a warning device enables pilots to concentrate on tasks which require their visual attention. In a study to determine the effectiveness of voice warning systems (Simpson & Williams, 1980), pilots were presented the task of recognizing a warning, recalling the immediate action, and determining the type of aircraft system which would be involved. The warnings were repeated until the pilot responded. The environment in which this study was performed was meant to reproduce as closely as possible that of an actual aircraft. The noise level was 75dB and cockpit conversation was allowed.

The pilots were presented with four different warning conditions: semantic context (landing gear not down), keyword context (gear not down), prior tone, no prior tone. The results of the study indicated that no alerting tone should occur before the voice warnings if synthesized voice is used only for warnings. However, if synthesized speech is used

elsewhere, as in feedback, it must be emphasized that the voice heard is a warning. The pilots also preferred the semantic context, noting that the longer messages were more informative. More information can be presented with only a slight increase in warning length and no effect on comprehension time.

In a study performed at the Air Force Institute of Technology (Freedman & Rumbaugh, 1983), the effects of background noise, signal-to-noise ratio, and type and length of precursor (warning before actual message) were analyzed. The background noise levels tested were 105dB and 115dB from a fighter aircraft. The signal to noise ratio levels tested were 0,5, and 10dB. Three different precursor formats were used: tone, voice, and repeated warnings. Accuracy and response time were measured.

The results of this study are:

a. Background Noise: Greater accuracy was achieved at 105dB, and there was no difference in the time of response.

b. Signal-to-Noise Ratio: The most accurate response was 10dB; there was no difference in the time of response.

c. Precursor: A repeated warning was associated with highest accuracy, and the tone precursor had the fastest response time. A repeated warning had the fastest response time after an adjustment was made to balance out the difference in the length of the warning between the different types of voices.

From this study, it was suggested that a repeated warning, where the warning acted as the attention-getting device, should be used.

Voice generation technology presently exists which would permit the utilization of voice warning systems. If the guidelines are properly followed, voice warning systems permit the pilots to concentrate on tasks requiring their visual attention. The pilots' workload is reduced and performance increases.

### Male, Female, or Machine Voice

Voice generation can be used to provide feedback to users of voice recognition systems or in warning systems. When the decision to utilize voice generation technology in a certain situation is made, it must then be decided which type of voice to use.

In a study conducted by Sikorsky Aircraft (Bertone, 1982), the author concluded that the female voice was superior to a male voice. Pilots must monitor two to four radio frequencies simultaneously. A female voice is distinctive and draws attention to the warning or feedback. Also, the female voice has a higher frequency than the male voice. Helicopter noise tends to diminish in the higher frequencies. Therefore, a higher frequency voice would be heard more easily than a lower frequency voice.

Freedman and Rumbaugh (1983) tested three types of voice generation to determine which had better accuracy and response time. Based upon the time of response, the female voice warning was associated with a faster response time compared to both the male and machine voices. However, the authors noted that the mean length of the female voice warning was shorter than the male voice warning and the machine voice warning. Believing that the faster response time was a result of the short message time, the authors adjusted the response time for the length of the warnings. After this adjustment, it was determined that the male voice warning was associated with the fastest response time, as well as the greatest accuracy.

Grumman Aerospace Corporation performed an experiment to determine the capability of recognizing and understanding synthesized female voice messages in the presence of human male voice radio messages, and evaluating the effect of the method of presentation of the radio, caution, and warning messages on message discrimination and intelligibility (Manaker, 1982). Subjects performed an aircraft tracking task while listening to tapes of air traffic control radio messages and a synthetic female voice caution and warning system.

Messages were presented in three ways:

a. Radio, caution, and warning messages were presented to both ears simultaneously.

b. Cautions and warnings were presented to one ear, while radio messages were presented to both ears.

c. Radio messages were presented to both ears. When a caution or warning message was presented, the radio message was not given to the ear receiving the caution or warning message.

Results of this experiment support that voice messages are effective. In this case, the female voice could be heard and understood even when radio messages were present. Also, important radio messages could be heard and understood while the synthetic female voice was presenting caution and warning messages. Manaker determined that there exists a difference in the ability of subjects to hear and understand radio messages for the different methods of presenting competing synthetic female voice caution and warning messages.

Since conclusive data do not exist to support the use of one certain type of voice, further studies should be conducted to pinpoint the advantages and disadvantages of each voice in different settings.

## CONCLUSIONS

The field of voice recognition and generation has advanced to where it has potential for use in future helicopter cockpits. The ability to enter data and request needed information by voice enables the pilot to maintain both hands on the flight controls. Also, the use of voice generation to relay the requested information back to the pilot and to present caution and warning messages relieves the pilots of the necessity to glance from their forward field of sight to the display panel to receive information.

As far as the question of whether voice data entry is more effective than manual data entry, the reviewed literature shows that, when the only task is to enter data points, manual entry is faster than voice entry. However, the voice response mode is more effective for complex data entry tasks. When required to perform a secondary task, such as pursuit tracking, the studies presented in this report have shown voice entry to be less disruptive than manual entry. Related to actual flight, these results would indicate that voice entry would have less impact on a pilot's flight performance during low-level flight and other difficult missions.

Vocabulary considerations are an important factor in the use of voice recognition. Vocabulary sizes should be limited to include only the necessary words for operation of the voice system. The selection of the actual vocabulary should be performed carefully to avoid the use of similar sounding words and words which could be confused in meaning. The use of improved syntax structures can help in increasing recognition performance.

Pilots are faced with an extremely high workload. The stress which results from the numerous tasks which must be performed affects the pilot's voice, affecting recognition performance. The studies reviewed in this report show that as mental loading, and the stress which results, increases, recognition accuracy decreases.

The noise levels experienced in a helicopter cockpit present another obstacle to be overcome by both voice recognition and synthesis. The use of noise-cancelling devices helps to alleviate the negative effects on recognition. The effects of noise on recognition can also be resolved by training the voice recognition system in the noise environment in which it is to be operated. Voice generation systems must be loud enough and of such a frequency that one can hear and understand the message being transmitted.

The application of voice generation systems presents less of a problem than voice recognition systems. Currently, voice warning systems are being used in a number of areas. The research performed dealing with voice warning systems suggests that repeated warnings should be used to present the caution and warning messages.

Another use for voice generation in the cockpit is as a form of feedback to the pilot. The studies performed to determine whether aural feedback is more effective than visual feedback suggest the use of voice feedback when other tasks requiring visual attention are present. However, it was also shown that voice feedback slows voice data entry. Therefore, when no secondary tasks are present, visual feedback should be used.

Based upon the research presented in this review, it seems feasible that voice recognition and generation can be effectively applied in future helicopter cockpits.

## RECOMMENDATIONS

The following areas are recommended for future research:

a. Stress and Workload - The effects of stress and increased workload are a major concern when considering the use of voice recognition in a helicopter environment. Research should deal with the effects of stress on the voice and the resulting effect on recognition performance. Also, the effect of an increased workload on the voice and recognition performance should be investigated further.

b. Vocabulary - Improved syntax structures should be developed to improve isolated word recognizer performances.

c. Voice Generation - The issue of which type of voice generation is best for use in helicopter cockpits should be investigated further.

d. Specific applications of voice recognition and generation should be investigated and tested for effectiveness.

## REFERENCES


1. Aretz, A. J. (1983). Comparison of manual and vocal response modes for the control of aircraft subsystems (Report No. AFWAL-TR-83-3005). Wright-Patterson Air Force Base, Ohio: Air Force Wright Aeronautical Laboratories-Flight Dynamics Laboratory.

2. Armstrong, J. W. (1980). Effects of concurrent motor tasking on performance of a voice recognition system. Monterey, CA: Naval Postgraduate School, Department of Operations Research. (ADA093557)

3. Armstrong, J. W., & Poock, G. K. (1981a). Effect of operator mental loading on voice recognition system performance. Monterey, CA: Naval Postgraduate School, Department of Operations Research. (ADA107477)

4. Armstrong, J. W., & Poock, G. K. (1981b). Effect of task duration on voice recognition system performance. Monterey, CA: Naval Postgraduate School, Department of Operations Research. (ADA107442)

5. Beek, B., & Vonusa, R. S. (1983). General review of military applications of voice processing. In J. Bridle (Director), AGARD (pp. 1.1-1.20). Essex, England: Specialised Printing Services Limited. (ADA132092)

6. Berson, B. L., Po-Chedley, D. A., Boucek, G. P., Hanson, D. C., & Leffler, M. F. (1981). Aircraft alerting systems standardization study: Aircraft alerting systems design guidelines (Report No. DOT/FAA/RD-81/38/II). Seattle, WA: Systems Technology, Research-Crew Systems, Boeing Commerical Airplane Co. (ADA106732)

7. Bertone, C. M. (1982). Human factors considerations in the development of a voice warning system for helicopters. Proceedings of the 1982 SAE Aerospace Congress & Exposition (pp.127-132).

8. Bridle, J. S. (1983). Connected-word recognition for use in military systems (Report No. AC-243-D/854). France: Defense Research Group, North Atlantic Council. (ADB072070).

9. Coler, C. R. (1983). Automated speech recognition for helicopter applications. Proceedings of 1983 American Voice Input/Output Society Conference.

10. Elster, R. S. (1980). The effects of certain background noises on the performance of a voice recognition system (Report No. NPS54-80-010). Monterey, CA: Naval Postgraduate School. (ADA106138)

11. Feuge, R. L., & Geer, C. W. (1978). Integrated applications of automated speech technology (Report No. ONR-CR213-158-1AF). Seattle, WA: Boeing Aerospace Company, Logistics Support and Services. (ADA053189)

12. Finkelstein, S. (1983). Customized LPC vocabulary lets cars talk. Speech Technology, 2, 65-69.

13. Freedman, J., & Rumbaugh, W. A. (1983). Accuracy and speed of response to different voice types in a cockpit voice warning system (Report No. LSSR 89-83). Wright-Patterson Air Force Base, OH: Air Force Institute of Technology, School of Systems and Logistics.

14. French, B. A. (1983). Some effects of stress on users of a voice recognition system: A preliminary inquiry. Monterey, CA: Naval Postgraduate School, Command, Control and Communications.

15. Groner, G. F., & Gilblom, D. L. (1983). Speech synthesis in industrial applications. Proceedings of 1983 American Voice Input/Output Society Conference.

16. Hogan, R., & Hogan, J. C. (1982). Subjective correlates of stress and human performance. In E. A. Alluisi & E. A. Fleisham (Eds.), Human performance and productivity: Stress and performance effectiveness (pp. 142-143). Hillsdale, NJ: Lawrence Erlbaum Associates.

17. Jay, G. T. (1981). An experiemnt in voice data entry for imagery interpretation reporting. Monterey, CA: Naval Postgraduate School. (ADA101823)

18. Malkin, F. J. (1983). Effects on computer recognition of speech when speaking through protective masks (Technical Memorandum 7-83). Aberdeen Proving Ground, MD: US Army Human Engineering Laboratory. (ADA140204)

19. Manaker, E. (1982). Pilot ability to understand synthetic voice and radio voice when received simultaneously. (Report No. ACT-R-82-01). Bethpage, NY: Grumman Aerospace Corporation, System Engineering Department.

20. North, R., & Lea, W. A. (1982). Application of advanced speech technology in manned penetration bombers (Report No. AFWAL-TR-82-3004). Wright-Patterson Air Force Base, OH: Flight Dynamics Laboratory. (ADA119274).

21. Nusbaum, H. C., Schwab, E. C., & Pisoni, D. B. (1983). Perceptual evaluation of synthetic speech: Some constraints on the use of voice response systems. Proceedings of 1983 American Voice Input/Output Society Conference.

22. Poock, G. K. (1981). A longitudinal study of computer voice recognition performances & vocabulary size (Report No. NPS55-81-013). Monterey, CA: Naval Postgraduate School. (ADA102208)

23. Poock, G. K., Schwalm, N. D., & Roland, E. F. (1982). Use of voice recognition equipment with stenographer masks (Report No. NPS55-82-028). Monterey, CA: Naval Postgraduate School.

24. Poock, G. K., Schwalm, N. D., Roland, E. F., & Martin, B. J. (1982). Trying for speaker independence in the use of speaker dependent voice recognition equipment (Report No. NPS55-82-032). Monterey, CA: Naval Postgraduate School.

25. Poock, G. K., Roland, E. F., & Schwalm, N. D. (1983). Wearing army gas masks while talking to a voice recognition system (Report No. NPS55-83-005). Monterey, CA: Naval Postgraduate School.

26. Reed, L. W. (1982). Voice interactive systems technology avionics (VISTA) program. Fort Monmouth, NJ: US Army Avionics Research and Development Activity. (ADA117288)

27. Ruess, J. C. (1982). Investigation into air launch cruise missile (ALCM) flight information loading & display techniques during flex targeting procedure. Monterey, CA: Naval Postgraduate School, Command, Control, and Communications. (ADA115744)

28. Simpson, C. A., & Williams, D. H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. Human Factors Journal, 22, 319-330.

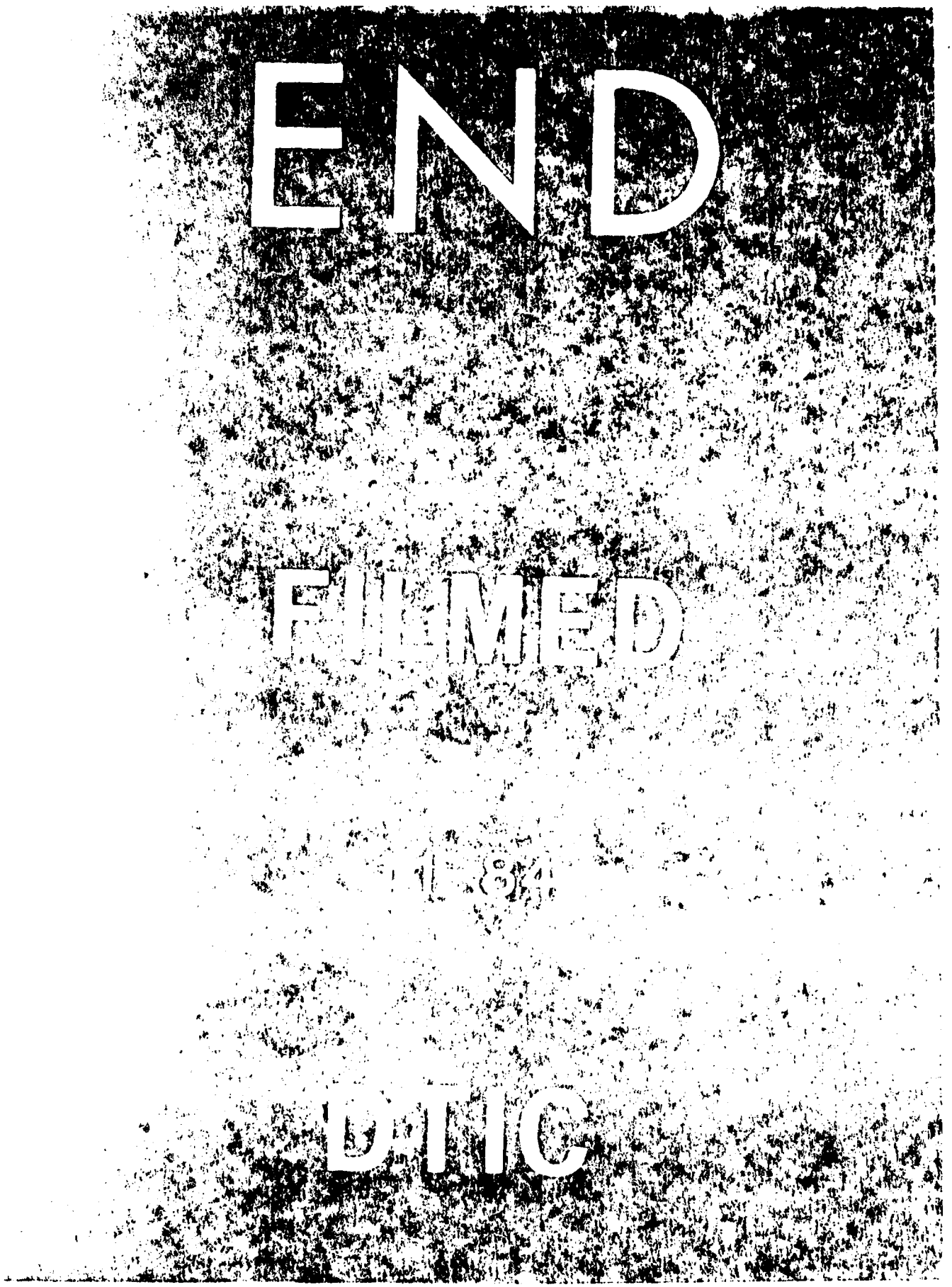
29. Skriver, C. P. (1979). Vocal & manual response modes: Comparison using a time-sharing paradigm (Report No. NADC-79127-60). Warminister, PA: Naval Air Development Center. (ADA119767)

30. Stokes, J. M., & Dow, L. (1980). Vocabulary development for the voice recognition and synthesis (VRAS) system (Report No. 1400.05-A). Willow Grove, PA: Analytics.

31. Voorhees, J. W., Marchionda, K. M., & Atchison, V. L. (1982). Speech command auditory display system (SCADS). Proceedings of 1982 International Conference on Acoustics, Speech & Signal Processing. (ADA117486)

32. Welch, J. R., & Shamsi, E. (1980). Advanced image exploitation aids (Report No. RADC-TR-80-74). Delran, NJ: Threshold Technology, Inc. (ADA085811)

33. Wyatt, B. (1983). Interactive voice control: Co-pilot of the future. Speech Technology, 2, 60-64.


END
FILMED
DTIC